

1 **Remarks for the “Response to Non-Final Office Action**
2 **Dated 2/17/2006”**

3 Applicant respectfully requests reconsideration and allowance of all of the
4 claims of the application. Claims 1-7, 15-19, 64, 65, and 67-71 are presently
5 pending. Claims amended herein are 1, 64, and 65. Claims withdrawn or
6 cancelled herein are none. New claims added herein are none.

7 **Telephone Conference**

8 Applicant appreciates the Examiner’s willingness to talk to the undersigned
9 attorney for the Applicant, Kasey Christie, on Thursday, March 23, 2006. The
10 focus of the conversation was on claim interpretation. In particular, we discussed
11 the interpretation of the following claim language (which is found in independent
12 claims 1, 15, 64, and 65): “obtaining a body of text containing textual content in a
13 computer-readable format.”

14 The Examiner asserted that virtually everything is “computer-readable”
15 because computers can process images acquired by scanners or digital cameras.
16 Of course, Applicant did not agree with that interpretation. Nevertheless,
17 Applicant amends the claim language in the independent claims 1, 64, and 65 to
18 clarify its interpretation.

19 Also, the Examiner inquired about the publication date of an article
20 referenced in an Information Disclosure Statement (i.e., form PTO-1449) dated
21 8/15/2003, which was authored by one of the co-inventors. The article is “R.
22 VENKATESAN, et al.: *Robust Image Hashing*, Cryptography Group, Microsoft
23 Research, 3 pages.” I contacted the inventor and he indicated that the article was
24 published on or after September of 2000.
25

In addition, the Examiner indicated that he would like a copy of the #10 reference from that article. That #10 reference is a manuscript that was authored by one of the co-inventors. The manuscript is "R. Venkatesan and S.-M. Koon, *Robust image hashing into binary strings*, 1999."

The inventor told me that he did not have a copy of this manuscript. He informed by that the manuscript was never published.

Substantive Claim Rejections

Claim Rejections under §103

The Office rejects all of the pending claims under §103. For the reasons set forth below, the Office has not shown that the cited references disclose, teach or suggest (under §103) the rejected claims. Accordingly, Applicant respectfully requests that the rejections be withdrawn and the case be passed along to issuance.

The Office's rejections are based upon **Hull: Hull et al.**, U.S. Patent No. 5,465,353 (issued 11/7/1995).

Overview of the Application

The Application describes a technology for recognizing the content of text documents. The technology may detect similarity between text-based works in an automatic and accurate manner. Furthermore, it may categorize content of text-based works in an automatic and accurate manner.

1 Generally, the technology determines one or more hash values for the
2 content of a text document. Furthermore, the technology may generate a "sifted
3 text" version of a document.

4 In one implementation described herein, document recognition is used to
5 determine whether the content of one document is copied (i.e., plagiarized) from
6 another document. This is done by comparing hash values of documents (or
7 alternatively their sifted text).

8 9 Hull

10 The Office cites **Hull** as its primary reference in its obviousness-based
11 rejections.

12 **Hull** discloses a document matching and retrieval system where an input
13 document is matched against a database of documents, using a descriptor database
14 which lists descriptors and points to a list of documents containing features from
15 which the descriptor is derived document. The descriptors are selected to be
16 invariant to distortions caused by digitizing the documents or differences between
17 the input document and its match in the document database. An array of
18 accumulators is used to accumulate votes for each document in the document
19 database as the descriptor base is scanned, wherein a vote is added to an
20 accumulator for a document if the document is on the list as having a descriptor
21 which is also found in the input document. The document which accumulates the
22 most votes is returned as the matching document, or the documents with more than
23 a threshold number of votes are returned
24
25

Obviousness Rejections Based upon Hull

The Office rejects all pending claims (claims 1-7, 15-19, 64-65 and 67-71) under USC § 103(a) as being unpatentable over **Hull**. Applicant respectfully traverses the rejections of these claims.

Claims 1, 64, 65

With this rejection, the Office has rejected the independent claims 1, 64, and 65. Applicant amends independent claims 1, 64, and 65 so that the first element reads in the following manner:

“obtaining a body of text containing textual content in a computer-readable format, wherein the textual content of the obtained computer-readable formatted body of text is mutable via software tools for manipulation of textual content of bodies of text.”

Applicant submits that this amendment clarifies the characteristics of the body of text which is the subject of the “obtaining” element of the claims. **Hull** does not disclose this.

Consequently, **Hull** does not disclose all of the claimed elements and features of these claims. Accordingly, Applicant asks the Office to withdraw its rejection of these claims.

Claims 2-7 and 67-69

These claims ultimately depend upon independent claim 1. As discussed above, claim 1 are allowable.

1 In addition to its own merits, each of these dependent claims is allowable
2 for the same reasons that its base claim is allowable. Applicant submits that the
3 Office withdraw the rejection of each of these dependent claims because its base
4 claim is allowable.

5 Furthermore, with regard to claim 67, the Office indicates the following on
6 page 6 of the Action:

7 With regards to the claim 67, Hull fails to explicitly disclose indicating suspicion of
8 plagiarism between the two bodies of text when the compared hash values of the two
9 bodies of text substantially match, as claimed. But, it is well known in the art to find out
10 the plagiarism using the hash comparison. Therefore, it would have been obvious to
11 one ordinary skilled in the art at the time of invention to simply incorporate the well
12 known feature of the comparison of hash values in order to detect the suspicion of
13 plagiarism in the text bodies.

14 Applicant respectfully disagrees with the Office's assertion that it is "well
15 known in the art to find out the plagiarism using the hash comparison." Indeed,
16 Applicant submits that hashes are not typically used to detect plagiarism because
17 of the desirable property of hashing functions for "apparently uniform
18 distribution" of the output. See p. 14 of the Application and
19 <http://en.wikipedia.org/wiki/Hashing>. Because of this property, one minor
20 difference between compared bodies of text is likely to result in largely disparate
21 hash values.
22

23 Applicant submits that one of ordinary skill in the art would not use
24 conventional hashing to detect plagiarism because of this uniform distribution of
25 output property of conventional hashing. If the use of hashing is well-known to

1 those in the field, then Applicant submits that the Office should submit a reference
2 that shows this.

3 Furthermore, Applicant submits that **Hull** does not disclose any suggestion,
4 teaching, or anything else to motivate a person of ordinary skill in the art to solve
5 any problems or issues related to plagiarism. **Hull** deals with the fields of image
6 processing and storage. Not plagiarism.

1 Claim 15

2 This claim reads, in part, as follows (with emphasis added):

- 3
- 4 • obtaining a body of text containing textual content in a computer-
 - 5 readable format;
 - 6 • formatting the body of text into a defined image-based format,
 - 7 wherein the textual content of the defined image-based formatted
 - 8 body of text is immutable via software tools for manipulation of
 - 9 textual content of bodies of text;
 - 10 • deriving a hash value representative of the body of text,
 - 11 **perceptually similar bodies of text having proximally similar**
 - 12 **hash values.**
 - 13

14 On page 5 of the Action, the Office indicates the following regarding this
15 claim:

16 With regards to Claims 15-17, arguments analogous to those presented for
17 Claims 1-5 are applicable to Claims 15-17. Hull further discloses similar bodies of text
18 having proximally similar hash values (Figures 2, numerical 214, and see col. 13 lines
19 30-44, wherein the hash descriptors are created using the word of the document and
20 the descriptors are "proximally similar hash values" as they are dependent on the
21 lengths of the words, and if the words are same the hash values will be proximally be
22 similar) as claimed.

Applicant respectfully submits that **Hull** does not disclose: “perceptually similar bodies of text having proximally similar hash values.” Rather, **Hull** discloses conventional hashing functions which possess the characteristic of producing output with an “apparently uniform distribution.” See p. 14 of the Application and <http://en.wikipedia.org/wiki/Hashing>.

Indeed, **Hull** describes this characteristic of the hash function that it discloses. At col. 10, lines 24-35, **Hull** discloses this (with emphasis added):

An example of hashing generator 214 will now be described. The operation of hashing generator 214 is best described by its “hash function” which maps a descriptor to a hashed descriptor. In a typical hashing operation each descriptor maps to a single hashed descriptor, i.e., it is a function, and there are fewer hashed descriptors than descriptors. **The primary concerns in the choice of hash function are to ensure a high rate of dispersion by the hash function** and to minimize the storage required for the data structures. High dispersion is that which results in a set of typical descriptors distributing fairly evenly over the hashed descriptors, a result of which is often referred to as having a low collision rate.

Applicant submits that a conventional hashing approach (which is disclosed by **Hull**) does not produce “proximally similar hash values” when given “perceptually similar bodies of text.”

Consequently, **Hull** does not disclose all of the claimed elements and features of this claim. Accordingly, Applicant asks the Office to withdraw its rejection of this claim.

Claims 16-19 and 70-71

These claims ultimately depend upon independent claim 15. As discussed above, claim 15 are allowable.

In addition to its own merits, each of these dependent claims is allowable for the same reasons that its base claim is allowable. Applicant submits that the Office withdraw the rejection of each of these dependent claims because its base claim is allowable.

1 **Dependent Claims**

2
3 In addition to its own merits, each dependent claim is allowable for the
4 same reasons that its base claim is allowable. Applicant submits that the Office
5 withdraw the rejection of each dependent claim where its base claim is allowable.
6

7 **Conclusion**

8 All pending claims are in condition for allowance. Applicant respectfully
9 requests reconsideration and prompt issuance of the application. If any issues
10 remain that prevent issuance of this application, the Office is urged to contact the
11 undersigned attorney before issuing a subsequent Action.
12

13 Respectfully Submitted,

14
15 Dated: 3.30.06

16 By: 

17 Kasey C. Christie
18 Reg. No. 40559
19 (509) 324-9256 x232
20 kasey@leehayes.com
21 www.leehayes.com
22
23
24
25